

# Secure Content Distribution in Vehicular Networks

Viet T. Nguyen, Jubin Jose, Xinzhou Wu and Tom Richardson

Qualcomm Research

Bridgewater, NJ

Email: {ntienvie, jjose, xinzhouw, tomr}@qti.qualcomm.com

**Abstract**—Dedicated short range communication (DSRC) relies on secure distribution to vehicles of a certificate revocation list (CRL) for enabling security protocols. CRL distribution utilizing vehicle-to-vehicle (V2V) communications is preferred to an infrastructure-only approach. One approach to V2V CRL distribution, using rateless coding at the source and forwarding at vehicle relays is vulnerable to a pollution attack in which a few malicious vehicles forward incorrect packets which then spread through the network leading to denial-of-service. This paper develops a new scheme called Precode-and-Hash that enables efficient packet verification before forwarding thereby preventing the pollution attack. In contrast to rateless codes, it utilizes a fixed low-rate precode and random selection of packets from the set of precoded packets. The fixed precode admits efficient hash verification of all encoded packets. Specifically, hashes are computed for all precoded packets and sent securely using signatures. We analyze the performance of the Precode-and-Hash scheme for a multi-hop line network and provide simulation results for several schemes in a more realistic vehicular model.

## I. INTRODUCTION

Security protocols designed for V2V communication [1] rely on the assumption of periodic distribution to all vehicles of a Certificate Revocation List (CRL) created by a certificate authority [2], [3], [4]. Since requiring every vehicle to have internet connectivity is a significant obstacle for large-scale adoption, it is important to develop an approach that requires only a few vehicles to obtain the CRL from infrastructure and utilizes V2V communication to distribute the CRL to the rest of the vehicles.

CRL distribution in vehicular networks is a typical file distribution problem with stringent security requirements. The approach of distributing packets in a round-robin or random fashion at the source and relaying (vehicles relaying packets to other vehicles) is very inefficient. Similar to the coupon collector problem, the inefficiency arises from the delay and redundancy increase for successive innovative packets. The utilization of an efficient fountain or rateless code at the source solves this issue. Furthermore, the original file can be appended with a signature to ensure that vehicles can detect any file alteration by malicious vehicles. However, an important and challenging security concern remains, which is pollution attack. The attack involves the forwarding of incorrect packets by one or more malicious vehicles. If these malicious packets are forwarded by other non-malicious relays then the incorrect packets will spread very fast and many vehicles will not be able to decode the original file, even when there is very few malicious nodes. Hence, secure content distribution requires that vehicles forward only verified packets.

If a signature is appended to the original file, then a relay can wait until it decodes the entire file to verify all the received packets. All vehicles (except possibly malicious vehicles) do not forward packets until the file is decoded. We refer to such a scheme as *Wait-to-Decode*. An important advantage is that re-encoding is possible after decoding. However, there are drawbacks: (a) If the file is large there could be a large initial delay before relays successfully decode and begin to forward. (b) Since incorrect packets can be forwarded by malicious vehicles to many vehicles, decoding at these vehicles could fail without error correcting methods.

Another technique to enable packet verification is to include signatures on each encoded packet after applying fountain coding at the source. The advantages of this *Sign-every-Packet* scheme include the ability to verify each individual packet and allow forwarding before decoding the file. However, there are disadvantages: (a) The overhead introduced is big and grows linearly with the number of packets in the file. (b) The computational complexity to verify each packet is high. (c) The relays cannot re-encode after the file is decoded as the signature can be included by only the source.

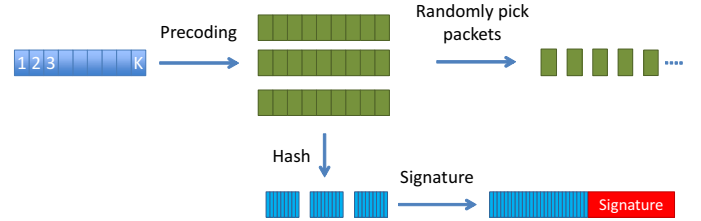


Fig. 1. Precode-and-Hash Scheme: A new scheme that enables efficient packet verification with low overhead

In this paper, we develop a new scheme, *Precode-and-Hash*, and characterize its performance using both analytical and simulation results. Although existing fountain codes are throughput efficient, adding packet verification into these schemes is challenging due to infinitely many potential coded packets. By using a simple precode (for example, of rate  $1/N$ ), we limit the number of potential coded packets to  $N$  times the file size at the expense of some throughput overhead. We take advantage of this to design an efficient packet verification scheme. Specifically, we separately hash all precoded packets and group them together into one or few hash-information packets. A fountain code is applied to these hash-information packets. The resulting coded hash-

information packets are signed by the source and distributed in the network by forwarding, possibly with higher priority in the beginning. This new scheme is captured in Figure 1. Once a relay has received these hash-information packets, it can verify each packet by taking its hash and comparing with the hash it received securely (i.e., hash-information packets that are signed). Since hashes are small (20 bytes for widely used SHA1 hash) and computationally efficient, packet verification is efficient both in terms of overhead and complexity. More details on Precode-and-Hash scheme can be found in Section III.

The main contributions in this paper include the following.

- We analyze our scheme for multi-hop line network and prove that a speed-up factor of at least 2 is possible over wait-to-decode scheme.
- We prove the convergence to fluid limit and show that a speed-up factor of  $e$  is achievable. Furthermore, we show that this result is tight.
- We provide detailed simulations of above mentioned schemes for Boston urban-area model in Section V.

Both analysis and simulations show that Precode-and-Hash is able to solve the pollution attack problem with limited overhead compared to a non-secure fountain code scheme. Hence, Precode-and-Hash is a highly desirable candidate for secure CRL distribution in vehicular networks.

#### A. Related Work

Network coding - a coding paradigm that allows re-encoding at relays - have the potential to achieve the maximum information flow capacity in a network [5], [6]. Network coding and its variants are, however, vulnerable to many security threats [7], one of which is the pollution attack. Several countermeasures have been proposed to cope with this challenge. The general approach is to design hash functions that allows hash of a linearly combined packet (produced by Network Coding) to be computed as a function of the hashes of the combining packets. Some examples include homomorphic hash [8], secure random checksum [9] and the very smooth homomorphic hash [10]. However, the complexity of computing these hashes is still very far from being practical. We, on the other hand, design an alternative delivery scheme that is sub-optimal, compared to network coding, but provides protection against pollution attack at very low computation cost using (any) standard hash functions.

## II. SYSTEM MODEL

A source (e.g. roadside unit) has a file (e.g. certificate revocation list) that consists of  $k$  packets. The source intends to send the same file to all the vehicles through a wirelessly connected (e.g. DSRC) vehicular network, where each transmission may be loss with probability  $\epsilon$ . A transmission is successful if the intended transmitter is within a certain range (transmission range) and all other transmitters are beyond a certain range (interference range) from the receiver. All the vehicles move around in a geographical region based on a mobility model.

We consider the following network topologies and mobility models.

- **Multi-hop line network:** This network is formed by a set of  $d$  static nodes placed along a line. Node 0 is the source while nodes  $1, 2, \dots, d-1$  are relays and destinations. Node  $i$  is in the range of nodes  $i-1$  and  $i+1$ . For example, node 0 is in range of node 1, node 1 is in range of nodes 0 and 2, and so on.
- **Boston urban-area model:** This model is expected to capture real-world topology and mobility. We model majority of the roads in Boston-Cambridge area and simulate traffic on this model. Each vehicle's movement at every intersection follows a Markov chain  $P = [P_{ij}]$ , where  $P_{ij}$  is the probability to switch from directed road segment  $i$  to directed road segment  $j$ . We calibrate  $P$  using the real daily traffic volume data reported for each major road segment in this area [11]. More details on this model are given in Section V.

## III. PRECODE-AND-HASH

The content distribution scheme is depicted in Figure (1). It consists of two main components:

- **Precode:** A fixed-rate (typically, low-rate) erasure code is applied as a precode. Let the rate of this precode be  $1/M$  (for example,  $1/3$  in Figure 1). If a file consists of  $k$  packets, there will be  $Mk$  coded packets in total. We assume an optimal erasure code so the file can be successfully decoded if a receiver has *any*  $k$  out of  $Mk$  coded packets.<sup>1</sup> Since the decoding is successful whenever a vehicle receives *any*  $k$  coded packets, the *rareness* issue is solved. More precisely, the  $\log k$  overhead arising in the coupon collector problem is reduced to a constant overhead, which decreases with lower code rate (or higher number of coded packets).
- **Hash:** Hashes are computed by the source for each coded packet using a sufficiently hard hash function. These hash packets must be distributed to the whole network before data packets are sent out. Since the number of hash-information packets is small, we expect the overhead for this initial distribution of hashes to be small. These hash make it possible for other nodes to quickly check if a coded packet is polluted.

A node, depending on its role, performs the following operations.

- A *source* chooses a random coded packet and broadcasts to its neighbors.
- A *relay* has to be both a receiver and a transmitter. As a transmitter, it chooses a random coded packet in the set of the coded packets that it receives and broadcast it to its neighbors. As a receiver, it discards all polluted packets and when its buffer contains at least  $k$  distinct coded packets, it can reconstruct the original file. Then,

<sup>1</sup>A practical code could introduce a small overhead, but it is often negligible with large file size.

that relay can apply the precode to the original file and acts as a secondary source.

#### IV. ANALYSIS OF DELAY PERFORMANCE

We are interested in the delay performance (i.e., the time required for all the nodes in the network to successfully receive the data file) for the precode-and-hash scheme. Since quantifying this delay seems intractable for finite file size, the focus is on the asymptotic delay, i.e., the limit of the ratio between the distribution delay and file size when the file size goes to infinity. This value is a good indicator of the overhead introduced by the Precode in the precode-and-hash scheme when the file size is sufficiently large.

##### A. Discreet Analysis

At time  $t$  some nodes have the file and the remainder have a subset  $H_i(t)$  of the coded packets. WLOG, we assume  $H_i(t) \subset H_{i-1}(t)$  and if node  $i$  has the file then so does node  $i-1$ . It is then sufficient to describe the system at time  $t$  by an infinite sequence of decreasing number  $\mathcal{H}(t) = \{|H_0(t)|, |H_1(t)|, \dots\}$ . It is easily seen that such a system is indeed Markovian, where  $|H_i(t+1)| = |H_i(t)| + 1$  with probability  $1 - \epsilon |H_i(t)| / |H_{i-1}(t)|$  and remains the same otherwise. It's also worth notice that when  $|H_i(t)| + 1 = k$ , that value is substituted by  $Mk$  and remains at this value as node  $i$  becomes a secondary source. Let  $T_n$  be the time  $t$  such that  $H_n(t)$  first become  $Mk$ , the first observation is encapsulated in the following Lemma.

**Lemma 1.**  $\mathbb{P}(|H_{n+1}(T_n)| / (k-1) \geq 0.5)$  converges to 1 as  $k \rightarrow \infty$ .

This Lemma shows that by the time node  $n-1$  become a secondary source, node  $n$  has already at least half the packets needed that are needed for decode. Hence the delivery time from node  $n-1$  to node  $n$  decreases even further, compared to the one hop delay from node 0 to node 1. The next Theorem characterizes this speed up.

**Theorem 1.** Assuming the relays initially have no packet, we have,

$$\mathbb{P}\left(\frac{(1-\epsilon)T_n}{k} \leq M \log \frac{M}{M-1} + (n-1)M \log \frac{2M-1}{2M-2}\right) \rightarrow 1$$

as  $k \rightarrow \infty$ . Hence,

$$\frac{(1-\epsilon)T_n}{k} \leq 1 + \frac{n-1}{2}$$

holds with large probability for large enough  $k, N$  and

$$\frac{(1-\epsilon)T_n}{nk} \leq M \log \frac{2M-1}{2M-2}$$

holds with large probability for large enough  $k, n$ .

So, the additional file decoding delay for multi-hop line network compared to single-hop grows as  $(n-1)/2$  ( $n$  is the number of hops) with Precode-and-Hash scheme, given the precode rate  $1/M$  is low enough. In contrast, the file decoding delay for Wait-to-Decode scheme grows as  $n-1$ .

##### B. Fluid limit Analysis

As the file size  $k \rightarrow \infty$ , we can prove that the file distribution converges sharply to the fluid limit (e.g. by Wormald's approach) for any finite time and length. The fluid limit model is as follows.

Let  $\frac{1}{k}|H_i(t)| = h_i(t) < 1$ . The function  $h_i(t)$  is non-decreasing in  $t$ , and continuous except at time  $\mathcal{T}_i$  when it reaches 1. At that point the node  $i$  is able to decode the file and then for  $t > \mathcal{T}_i$  we have  $h_i(t) = M$ . For  $i > 0$  and  $t \notin \{\mathcal{T}_{i-1}, \mathcal{T}_i\}$  we have  $\frac{d}{dt}h_i(t) = 1 - \frac{h_i(t)}{h_{i-1}(t)}$ . For  $t \in \{\mathcal{T}_{i-1}, \mathcal{T}_i\}$  we have  $\frac{d}{dt}h_i(t) = 1 - \frac{M}{h_{i-1}(t)}$ . Then, we can further improve the asymptotic bound on distribution delay in  $n$ -hop chain in Proposition 1. Furthermore, the later bound is tight. The main result of this analysis is given next.

**Theorem 2.** Assuming the initial condition  $h_i(0) = 0, i > 0$ , we have, for  $M \geq 2$ ,

$$\lim_{n \rightarrow \infty} \frac{\mathcal{T}_n}{n} = T^F(M)$$

where  $T^F(M)$  is the unique  $x \in [0, 1]$  solving  $-\ln x = 1 - \frac{x}{M}$ .

Note that the file size is scaled down to 1, hence we should have  $\mathcal{T}_i \approx \frac{(1-\epsilon)T_i}{k}$  (the term  $(1-\epsilon)$  accounts for lost packets). The above result can be rewritten as follows.

**Corollary 1.** For very large  $k, n$ , the followings hold with large probability

$$\frac{(1-\epsilon)T_n}{kn} = T^F(M) + o(1)$$

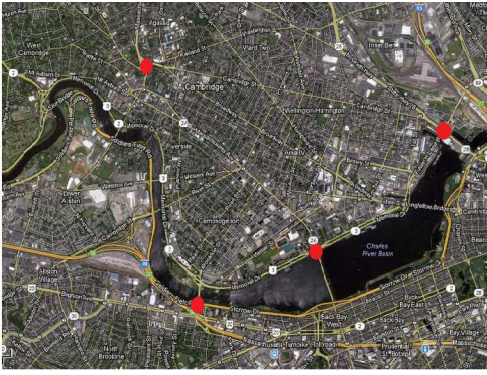
Compare to Proposition 1, the asymptotic bound decreases by a factor at least  $e/2$  and can be up to 1.74 (when  $M = 2$ ).

#### V. SIMULATION RESULTS & COMPARISON

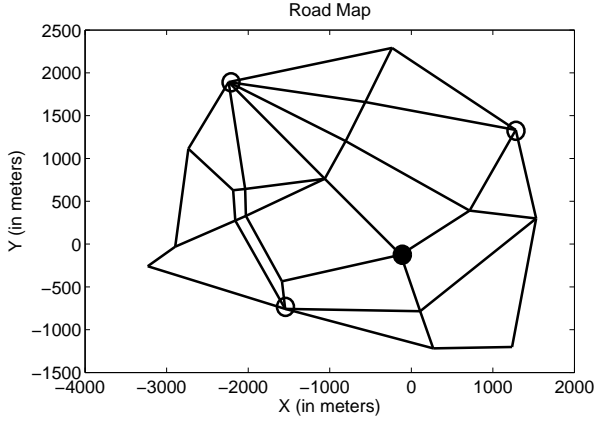
##### A. Scenarios & Parameters

The primary application of interest is CRL (certificate revocation list) distribution in vehicular networks. Each vehicle is securely identified by its pseudonym ID, which is issued by the Certification Authority (CA). For privacy reasons (e.g., preventing tracking), each vehicle is expected to change its pseudonym periodically (e.g., once every 10 minutes). Each vehicle is equipped with sufficient pseudonyms for long term operation. If the CA distrusts a vehicle, then it has to revoke all pseudonyms associated with that vehicle. This information is distributed to all other vehicles in a CRL file. CRL information corresponding to each vehicle is expected to be around 40 kB or higher. Hence, assuming tens of vehicles in the list, CRL file size is expected to be few MB. In our simulation, we assume that the CRL file size is 1 MB which is split into 1000 packets of 1000 bytes each.

We model V2V wireless communication using the following abstractions. The physical layer is modeled as an erasure channel with a packet erasure probability of  $\epsilon = 0.05$ . The medium access layer (MAC) is modeled by slotted carrier sensing (CSMA/CA). A transmission is successful if the intended transmitter is within a certain range (transmission



(a) Satellite map of Boston urban-area



(b) Road map of Boston urban-area

Fig. 2. Map of Boston urban-area

range of 200 m) and all other transmitters are beyond a certain range (interference range of 300 m) from the receiver. A vehicle can transmit 20 packets each time slot if it is elected by CSMA/CA.

We consider 236 vehicles moving with average velocity of  $20m/s$ , where each vehicle has a time-invariant velocity randomly (uniform) selected from  $[15, 25]m/s$ . There are four sources at fixed locations as shown in Figure 2. The sources continuously seed the file for a time period equal to 2 – 5 times the time to seed the given file size. For a fair comparison between different schemes, the seeding time is defined with respect to the original file size instead of the actual number of source packets. The seeding rate, i.e., the transmission rate of sources, is chosen to be 60 packets per second.

### B. Distribution Schemes and Security Overhead

We recall that the scheme developed in this paper for secure content distribution is Precode-and-Hash.

The hash-information packets are encoded and distributed using the Sign-every-Packet scheme. Each arriving packet is classified as, authentic, polluted or unknown (hash not received). Each vehicle is allowed to forward either a coded hash-information packet or an authentic coded data packet. Initially, a vehicle transmits coded hash-information packets for a fixed duration depending on the total size of the hashes.

After this initial phase, a vehicle sends coded hash-information packets with probability 0.2 and coded data packet with probability 0.8..

The intuition is that, a vehicle initially has more coded hash packets than coded data packets, so it will send coded hash packets first to ensure that the hashes are distributed “ahead” of the data. After that, we reserve enough capacity to distribute the hashes by choosing the right hash forwarding probability (0.2), which is fine-tuned by experiments.

As described in Section I, two other schemes that could provide protection against pollution attack are Wait-to-Decode and Sign-every-Packet.

The security overhead associated with each scheme is:

- Wait-to-Decode: A signature is appended to the whole CRL file. The signature size is 256 bytes for RSA-SHA2. Hence we can neglect this overhead.
- Sign-every-Packet: Each packet reserves 256 bytes for signature. The number of packets increases to  $\lceil 10^6 / (1000 - 256) \rceil = 1344$ . The overhead (in terms of packets) is about 34%.
- Precode-And-Hash: Each hash is 20 bytes in size. As each hash-information packets is individually signed after the fountain coded is applied, each packet contains at most  $\lfloor (1000 - 256) / 20 \rfloor = 37$  hashes and there are at least  $1000M / 37 \approx 27M$  such packets. The overhead is 8% with  $M = 3$ , 11% with  $M = 4$  and 13% with  $M = 5$ . In our simulations,  $M = 3$ .
- Genie Precode: This schemes use a rateless Precode instead of a finite rate one. It is considered here for comparison of data (excluding hash) delivery delay.

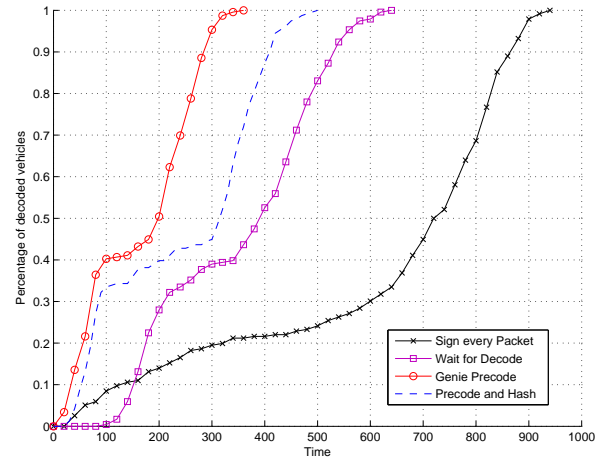


Fig. 3. Performance comparison: Boston urban-area model, seeding time enough to transmit 5 times the file size

### C. Simulation Results

There are four sources that seed the file. The seeding period equal to 5 times the duration to seed the file size (Figure 4). Next, to understand the performance of different schemes

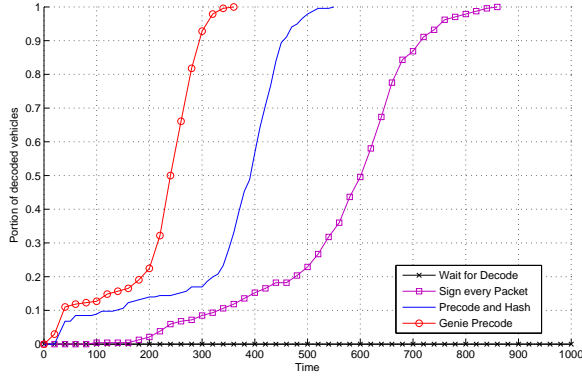


Fig. 4. Performance comparison: Boston urban-area model, seeding time enough to transmit 2 times the file size

with smaller seeding time, we reduce the seeding period to 2 (Figure 3). In all simulations, we plot the fraction of vehicles that has decoded the file as a function of time. Based on the result, the following remarks is made.

- Precode-and-Hash out-performs both Sign-every-Packet and Wait-for-Decode in the scenarios considered. Wait-for-Decode has very large file distribution delays.
- Precode-and-Hash is 20% to 30% worse than Genie Precode. This is expected since we reserve 20% capacity for hash distribution.
- In Figure 4, Wait-for-Decode distribution and Sign-every-Packet has a crossover due to smaller file size for Wait-for-Decode (recall that files size in Sign-every-Packet is 34% larger than other schemes). However, Sign-every-Packet is still better compared to Wait-for-Decode for majority of the vehicles.
- Both Wait-for-Decode and Sign-every-Packet are very sensitive to seeding parameters while Precode-and-Hash is more robust. In Wait-for-Decode, if no relay decodes file during the seeding period, no one can start re-seed the file after that. Hence, the file is not distributed at all. In Sign-every-Packet, as relays can only forward coded packets that they receive, only a small subset of coded packets can be distributed if the seeding time is not long enough. The decoding is sub-optimal as the innovative information is limited. In Figure 3, when the seeding time is reduced to 2, (a) Wait-for-Decode does not spread the file, (b) the file distribution time of Sign-every-Packet increases considerably while (c) the file distribution time of Precode and Hash only increases by 10%.

#### D. Practical Discussion

In the Precode and Hash scheme, the first practical choice one has to make is the Precode rate. This involves balancing the data delivery time and the capacity reserved for hash distribution as discussed in V-B. Note that the capacity reserved for hash distribution is much larger than the hash size itself since the hash distribution is much less efficient than data

distribution

Another observation is that the speed-up in file propagation, which can be up to  $e$  comes at the cost of an infinite ratio of wasteful transmissions (i.e., the transmitted packet is already in the receiver buffer). As in wireless network environment, extraneous transmissions increases the contention and interference (and hence increases packet loss probability) significantly, this problem may become very serious.

One way to avoid this issue is to let node forward packets with a probability proportioned to its buffer size. An analysis of this proportioned forwarding scheme is provided in Appendix G. The result shows that we still achieve a speed up factor at least 1.66 by using much less forwarding.

## VI. CONCLUSION

The Precode-and-Hash scheme developed in this paper provides protection against pollution attacks with a minimal increase in file delivery time. Hash verification is very efficient compared to signature verification. Since the file distribution delay is limited by physical mobility, the increase in delay of file distribution due to the use of finite-rate precode is minimal. However, the hash distribution delay can be significant. We expect that the net delay can be further improved with more optimized hash distribution. In summary, the Precode-and-Hash scheme is an attractive candidate for secure content distribution in vehicular networks.

## REFERENCES

- [1] "Standard specifications for telecommunications and information exchange between roadside and vehicle systems - 5 GHz band dedicated short range communications (DSRC) medium access control (MAC) and physical layer (PHY) specifications," Sept. 2003.
- [2] P. Papadimitratos, G. Mezzour, and J.-P. Hubaux, "Certificate revocation list distribution in vehicular communication system," *Proc. ACM International Workshop on Vehicular Internetworking*, San Francisco, CA, 2008.
- [3] K. P. Laberteaux, J. J. Haas, and Y.-C. Hu, "Security certificate revocation list distribution for VANET," *Proc. ACM International Workshop on Vehicular Internetworking*, San Francisco, CA, 2008.
- [4] M. Nowatowski and H. Owen, "Scalable certificate revocation list distribution in vehicular ad hoc networks," *Proc. SWiM'10*, Miami, FL, 2010.
- [5] R. Ahlswede, N. Cai, S.-Y. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.
- [6] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.
- [7] J. Dong, R. Curtmola, R. Sethi, and C. Nita-Rotaru, "Toward secure network coding in wireless networks: Threats and challenges," in *4th Workshop on Secure Network Protocols (NPSec)*, Oct 2008, pp. 33–38.
- [8] M. Krohn, M. Freedman, and D. Mazieres, "On-the-fly verification of rateless erasure codes for efficient content distribution," in *IEEE Symposium on Security and Privacy*, May 2004, pp. 226–240.
- [9] C. Gkantsidis and P. Rodriguez, "Cooperative security for network coding file distribution," in *IEEE International Conference on Computer Communications (INFOCOM)*, April 2006, pp. 1–13.
- [10] Q. Li, J.-S. Lui, and D.-M. Chiu, "On the security and efficiency of content distribution via network coding," *Dependable and Secure Computing, IEEE Transactions on*, vol. 9, no. 2, pp. 211–221, March 2012.
- [11] MassDOT, "Internet: www.mhd.state.ma.us," 2012.



## APPENDIX

### A. Proof of Proposition ??

From Lemma ?? we have

$$\begin{aligned}\mathbb{E}\left[\frac{t_{1,k}}{k}\right] &= \frac{1}{k} \sum_{i=0}^{k-1} \mathbb{E}[t_{1,i+1} - t_{1,i}] = \sum_{i=0}^{k-1} \frac{N}{(kN-i)(1-\epsilon)} \\ &= \frac{N}{1-\epsilon} \frac{1}{kN} \sum_{i=0}^{k-1} \frac{1}{(1-\frac{i}{kN})} = \frac{N}{1-\epsilon} \int_0^{1/N} \frac{1}{1-t} dt + O(k^{-1}) \\ &= \frac{N}{1-\epsilon} \log\left(\frac{N}{N-1}\right) + O(k^{-1}).\end{aligned}$$

To show convergence in probability we show that the variance of  $\frac{t_{1,k}}{k}$  tends to 0. Now, since  $t_{1,i+1} - t_{1,i}$  are independent random variables,

$$\begin{aligned}\text{Var}(t_{1,k}) &= \sum_{i=0}^{k-1} \text{Var}(t_{1,i+1} - t_{1,i}) = \sum_{i=0}^{k-1} p_i^{-2}(1-p_i) \\ &\leq \sum_{i=0}^{k-1} p_i^{-2} = \frac{kN}{(1-\epsilon)^2} \frac{1}{kN} \sum_{i=0}^{k-1} \frac{1}{(1-\frac{i}{kN})^2} \\ &= \frac{kN}{(1-\epsilon)^2} \int_0^{\frac{1}{N}} \frac{1}{(1-t)^2} dt + o(1) \\ &= \frac{kN}{(1-\epsilon)^2(N-1)} + o(1).\end{aligned}$$

It now follows, e.g. from Chebyshev's inequality, that  $\frac{t_{1,k}}{k}$  converges in probability to  $\frac{N}{1-\epsilon} \log\left(\frac{N}{N-1}\right)$ . We use the fact that  $\log(1+x) \leq x$  for  $x > -1$  for the bound.

### B. Proof of Subsection IV-A

The following lemmas are needed in the proof of Proposition 1.

First, we need to characterize the distribution of  $I_{n,i}$ . Let us introduce the notation  $\partial t_{n,i} = t_{n,i+1} - t_{n,i}$  and  $\mathcal{T}_{n,k} = (t_{n,1}, \dots, t_{n,k})$ .

**Lemma 2.** *Conditioned on  $\mathcal{T}_{n-1,k}$ , the r.v.s  $I_{n,i}$ ,  $i = 1, \dots, k$ , are  $k$  pairwise negatively dependent  $\{0,1\}$ -values r.v.s with conditional distribution*

$$\mathbb{P}(I_{n,i} = 0 \mid \mathcal{T}_{n-1,k}) = \prod_{j=i}^{k-1} \left(1 - \frac{1-\epsilon}{j}\right)^{\partial t_{n-1,j}}.$$

*Proof:* By definition, packet  $i$  is available at node  $n-1$  starting from time  $t_{n-1,i}$ . For each subsequent time the probability that packet  $i$  is not transmitted to the second hop is  $1 - \frac{1-\epsilon}{j}$ , hence the stated distribution holds.

First we note that for any finite random variables  $X, Y$   $\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(1-X)(1-Y) - \mathbb{E}(1-X)\mathbb{E}(1-Y)$  so we show negative dependence of  $\{1 - I_{n,i}, i = 1, \dots, k\}$ . Let  $i_1 < i_2$ , then

$$\begin{aligned}\mathbb{E}[(1 - I_{n,i_1})(1 - I_{n,i_2}) \mid \mathcal{T}_{n-1,k}] \\ = \mathbb{P}(I_{n,i_1} = 0 \text{ and } I_{n,i_2} = 0 \mid \mathcal{T}_{n-1,k}).\end{aligned}$$

The above expression is the probability that neither  $i_1$  nor  $i_2$  are transmitted to the second hop at any time from 1 to  $t_{n-1,k}$ .

Packet  $i_1$  is available from time  $i_1$  and both packets  $i_1$  and  $i_2$  are available from time  $i_2$ . After this time, the probability that neither is successfully transmitted in a given slot while  $j$  packets are available is  $1 - 2\frac{1-\epsilon}{j} \leq (1 - \frac{1-\epsilon}{j})^2$ . Thus we obtain

$$\begin{aligned}\mathbb{P}(I_{n,i_1} = 0 \text{ and } I_{n,i_2} = 0 \mid \mathcal{T}_{n-1,k}) \\ = \prod_{j=i_1}^{i_2-1} \left(1 - \frac{1-\epsilon}{j}\right)^{\partial t_{n-1,j}} \prod_{j=i_2}^{k-1} \left(1 - 2\frac{1-\epsilon}{j}\right)^{\partial t_{n-1,j}} \\ \leq \prod_{j=i_1}^{i_2-1} \left(1 - \frac{1-\epsilon}{j}\right)^{\partial t_{n-1,j}} \prod_{j=i_2}^{k-1} \left(1 - \frac{1-\epsilon}{j}\right)^{2(\partial t_{n-1,j})} \\ = \prod_{j=i_1}^{k-1} \left(1 - \frac{1-\epsilon}{j}\right)^{\partial t_{n-1,j}} \prod_{j=i_2}^{k-1} \left(1 - \frac{1-\epsilon}{j}\right)^{\partial t_{n-1,j}} \\ = \mathbb{P}(I_{n,i_1} = 0 \mid \mathcal{T}_{n-1,k}) \mathbb{P}(I_{n,i_2} = 0 \mid \mathcal{T}_{n-1,k}). \text{ a.s.}\end{aligned}$$

This completes the proof.  $\blacksquare$

The next step is to characterize the distribution of  $t_{n-1,1}, \dots, t_{n-1,k}$ .

**Lemma 3.** *For any  $k$ , there exist  $k-1$  independent geometric r.v.s  $\tau_1, \dots, \tau_{k-1}$  of parameters  $\frac{(1-\epsilon)(kN-1)}{kN}, \dots, \frac{(1-\epsilon)(kN-k+1)}{kN}$  respectively such that  $t_{n,j+1} - t_{n,j} \geq \tau_j$  a.s. for  $j = 1, \dots, k-1$ .*

*Proof:* Let  $\mathcal{E} := \{1, \dots, k\}^n$ . The  $k$ -tuple comprised of the number of packets collected by the  $k$  hops at each time slot form a Markov chain with state space  $\mathcal{E}$ . The fact that  $t_{k,j+1} - t_{k,j}$ ,  $j = 1, \dots, M-1$  are independent comes from the strong Markov property and from the fact that  $t_{n,j}$ ,  $j = 1, \dots, k$  are  $k$  increasing stopping times of this Markov chain. To complete the proof, we now need to show that

$$\mathbb{P}(t_{n,j+1} - t_{n,j} > l) \geq \left(1 - \frac{(1-\epsilon)(kN-j)}{kN}\right)^l.$$

In particular, let  $H_{t_{n,j}}, H_{t_{n,j}+1}, \dots, H_{t_{n,j}+l-1}$  be the number of coded packets at the  $n-1$  hop at time  $t_{n,j}, \dots, t_{n,j}+l-1$ . Note that the set of packets received at node  $n$  at any time must be a subset of the set of packets received at node  $n-1$ . At slot  $t_{n,j}+i$ , a new packet is received at node  $n$  iff node  $n-1$  chooses one of the  $H_{t_{n,j}+i} - j$  that node  $n$  does not have and the transmission is successful. The probability of this event is  $\frac{(H_{t_{n,j}+i}-j)(1-\epsilon)}{H_{t_{n,j}+i}}$ . So,

$$\begin{aligned}\mathbb{P}(t_{n,j+1} - t_{n,j} > l \mid H_{t_{n,j}}, H_{t_{n,j}+1}, \dots, H_{t_{n,j}+l-1}) \\ = \prod_{i=0}^{l-1} \left(1 - \frac{(H_{t_{n,j}+i}-j)(1-\epsilon)}{H_{t_{n,j}+i}}\right).\end{aligned}$$

On the other hand  $H_{t_{n,j}+i} \leq kN$  a.s. for  $i = \overline{0, l-1}$  a.s. Hence,

$$\begin{aligned}\mathbb{P}(t_{n,j+1} - t_{n,j} > l) \\ = \mathbb{E}[\mathbb{P}(t_{n,j+1} - t_{n,j} > l \mid H_{t_{n,j}}, H_{t_{n,j}+1}, \dots, H_{t_{n,j}+l-1})] \\ \geq \left(1 - \frac{(kN-j)(1-\epsilon)}{kN}\right)^l.\end{aligned}$$

This completes the proof.  $\blacksquare$

Two r.v.s  $X$  and  $Y$  are negatively dependent iff  $\mathbb{E}[XY] \leq \mathbb{E}[X]\mathbb{E}[Y]$ .

**Lemma 4.** Two r.v.s  $X, Y$  are negatively dependent if  $c - X, c - Y$  are negatively dependent, where  $c$  is any real number.

*Proof:* Since  $c - X, c - Y$  are negatively dependent,

$$\mathbb{E}[(c - X)(c - Y)] \leq \mathbb{E}[c - X]\mathbb{E}[c - Y].$$

Expanding the left hand side and the right hand side gives,

$$\begin{aligned} \mathbb{E}[c - X]\mathbb{E}[c - Y] &= (c - \mathbb{E}[X])(c - \mathbb{E}[Y]) \\ &= c^2 - c\mathbb{E}[X] - c\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ \mathbb{E}[(c - X)(c - Y)] &= \mathbb{E}[c^2 - cX - cY + XY] \\ &= c^2 - c\mathbb{E}[X] - c\mathbb{E}[Y] + \mathbb{E}[XY]. \end{aligned}$$

This implies

$$\mathbb{E}[XY] \leq \mathbb{E}[X]\mathbb{E}[Y].$$

Hence,  $X, Y$  are negatively dependent.  $\blacksquare$

The following lemma bounds the variance of the sum of mutually negatively dependent random variables (r.v.s).

**Lemma 5.** Let  $X_1, \dots, X_n$  be  $n$  mutually negatively dependent r.v.s. Then,

$$\text{Var}\left(\sum_{i=1}^n X_i\right) \leq \sum_{i=1}^n \text{Var}(X_i).$$

*Proof:* First of all,

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^2\right] &= \mathbb{E}\left[\sum_{i=1}^n X_i^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j\right] \\ &= \sum_{i=1}^n \mathbb{E}[X_i^2] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i X_j] \\ &\leq \sum_{i=1}^n \mathbb{E}[X_i^2] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i] \mathbb{E}[X_j]. \end{aligned}$$

Moreover,

$$\left(\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)\right]\right)^2 = \sum_{i=1}^n (\mathbb{E}[X_i])^2 + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i] \mathbb{E}[X_j].$$

This leads directly to the result.  $\blacksquare$

Now we can process to the main proof of Proposition 1.

*Proof:* We start by computing  $\mathbb{E}[N_n]$ . By Lemma 2,

$$\begin{aligned} \mathbb{E}[N_n] &= \sum_{i=1}^{k-1} \mathbb{E}[I_{n,i}] \\ &= (k-1) - \sum_{i=1}^{k-1} \mathbb{E}\left[\prod_{j=i}^{k-1} \left(1 - \frac{1-\epsilon}{j}\right)^{t_{n-1,j+1}-t_{n-1,j}}\right]. \end{aligned}$$

By Lemma 3,  $t_{n-1,j+1} - t_{n-1,j}$ ,  $j = 1, \dots, k-1$  are bounded below a.s. by  $k-1$  independent geometric r.v.s  $\tau_j$ ,  $j =$

$1, \dots, k-1$  with parameters  $p_j = (1-\epsilon)(kN-j)/kN$ ,  $j = 1, \dots, k-1$  respectively. Hence,

$$\begin{aligned} &\mathbb{E}\left[\prod_{j=i}^{k-1} \left(1 - \frac{1-\epsilon}{j}\right)^{t_{n-1,j+1}-t_{n-1,j}}\right] \\ &= \prod_{j=i}^{k-1} \mathbb{E}\left[\left(1 - \frac{1-\epsilon}{j}\right)^{t_{n-1,j+1}-t_{n-1,j}}\right] \\ &\leq \prod_{j=i}^{k-1} \mathbb{E}\left[\left(1 - \frac{1-\epsilon}{j}\right)^{\tau_j}\right] \end{aligned}$$

Now for any positive constant  $c \leq 1$  we have  $\mathbb{E}(c^{\tau_j}) = \sum_{k=1}^{\infty} c^k (1-p_j)^{k-1} p_j = \frac{cp_j}{1-c+cp_j}$ . Setting  $c = 1 - \frac{1-\epsilon}{j}$  we have  $cp_j < 1 - \epsilon$  so we obtain  $\frac{cp_j}{1-c+cp_j} \leq \frac{1}{\frac{1}{j}+1} = \frac{j}{j+1}$ . Continuing, we now have

$$\begin{aligned} \mathbb{E}[N_n] &\geq k-1 - \left(\sum_{i=1}^{k-1} \prod_{j=i}^{k-1} \frac{j}{j+1}\right) = k-1 - \left(\sum_{i=1}^{k-1} \frac{i}{k}\right) \\ &= k-1 - \frac{(k-1)k}{2k} = \frac{k-1}{2}. \end{aligned}$$

Next, we bound  $\text{Var}(N_n)$ . Using the fact that  $I_{n,i}$ ,  $i = 1, \dots, k-1$  have negative pair-wise dependence (Lemma 2) conditioned on  $t_i$ ,  $i = 1, \dots, k-1$ , we obtain

$$\begin{aligned} \text{Var}(N_n) &= \mathbb{E}[\text{Var}(N_n | t_{n-1,1}, \dots, t_{n-1,k-1})] \\ &\leq \mathbb{E}\left[\sum_{i=1}^k \text{Var}(I_{n,i} | t_{n-1,1}, \dots, t_{n-1,k-1})\right] \\ &= \sum_{i=1}^k \text{Var}(I_{n,i}). \end{aligned}$$

Since  $I_{n,i}$  are  $\{0, 1\}$  r.v.s,  $\text{Var}(I_{n,i}) \leq 1$ ,  $\text{Var}(N_n) \leq k$ . Then, by Chebyshev's inequality, we have

$$\begin{aligned} \mathbb{P}\left(\frac{1}{2} - \frac{N_n}{(k-1)} > \delta\right) &= \mathbb{P}\left(\frac{(k-1)}{2} - N_n > \delta(k-1)\right) \\ &\leq \mathbb{P}(\mathbb{E}[N_n] - N_n > \delta(k-1)) \\ &\leq \mathbb{P}(|\mathbb{E}[N_n] - N_n| > \delta(k-1)) \\ &\leq \frac{\text{Var}(N_n)}{\delta^2(k-1)^2} \leq \frac{k}{\delta^2(k-1)^2}. \end{aligned}$$

This completes the proof.  $\blacksquare$

*Proof of Proposition 1*

*Proof:* We use induction on  $n$ . Note that the  $n^{\text{th}}$  hop is the  $n+1^{\text{th}}$  node. The base case  $n=1$  is covered by Subsection ???. Suppose that the result holds for  $n-1$ . We can write  $T_n = T_{n-1} + \Delta$ , where  $T_{n-1}$  is the time it takes for node  $n$  to collect  $k$  coded packets. By the induction hypothesis,

$$\mathbb{P}\left(\frac{(1-\epsilon)T_{n-1}}{k} \leq N \log \frac{N}{N-1} + (n-2)N \log \frac{2N-1}{2N-2}\right)$$

converges to 1 as  $k$  goes to  $\infty$ . Moreover, by Proposition 1,  $\mathbb{P}(N_n/(k-1) \geq 0.5)$  converges to 1. So, by an argument

similar to that in Subsection ??,  $(1 - \epsilon)\Delta/k$  converges to

$$\begin{aligned} \frac{\sum_{i=N_n+1}^k \frac{kN}{kN-i}}{k} &= N \int_{N_n/kN}^{\frac{1}{N}} \frac{1}{1-x} dx + o(k^{-1}) \\ &= N \left( -\log \left( 1 - \frac{1}{N} \right) + \log \left( 1 - \frac{N_n}{kN} \right) \right) + o(k^{-1}) \end{aligned}$$

in probability a.s. (conditioned on  $N_n$ ). In the above equality, by substituting  $N_n$  by  $k/2$ , we get

$$\begin{aligned} &N \left( -\log \left( 1 - \frac{1}{N} \right) + \log \left( 1 - \frac{1}{2N} \right) \right) \\ &= N \left( -\log \left( \frac{N-1}{N} \right) + \log \left( \frac{2N-1}{2N} \right) \right) \\ &= N \log \left( \frac{2N-1}{2N} \frac{N}{N-1} \right) = \log \left( \frac{2N-1}{2N-2} \right). \end{aligned}$$

Hence,

$$\lim_{k \rightarrow \infty} \mathbb{P} \left( (1 - \epsilon) \frac{\Delta}{k} \leq N \log \frac{2N-1}{2N-2} \right) = \lim_{k \rightarrow \infty} \mathbb{P} \left( \frac{N_n}{k} \geq \frac{1}{2} \right) = 1.$$

The proof follows from this.  $\blacksquare$

### C. Proof of Theorem 2

An ancillary result that we will use repeatedly without further explanation is the following. Consider the differential equation

$$\frac{d}{dt}x(t) = -a(t)x(t) + b(t)$$

with  $x(0) = x_0 \geq 0$ ,  $a(t), b(t)$  Lipschitz continuous,  $a(t) > 0, b(t) \geq 0$  and  $a(t)$  bounded above. Then  $x(t) \geq 0$ . This follows from standard existence and uniqueness results on differential equations [?] and on the solution using Duhamels principle:

$$x(t) = x_0 e^{-z(t)} + \int_0^{z(t)} e^{-z(t)+s} \frac{b(z^{-1}(s))}{a(z^{-1}(s))} ds$$

where  $z(t) = \int_0^t a(u) du$ .

**Lemma 6 (Monotonicity).** *If  $h_i(0) \geq \tilde{h}_i(0)$  for all  $i$  then  $h_i(t) \geq \tilde{h}_i(t)$  for all  $i$  and  $t$ .*

*Proof:* Assume  $h_i(0) \geq \tilde{h}_i(0)$ . It follows that  $T_1 \leq \tilde{T}_1$  and that  $h_1(t) \geq \tilde{h}_1(t)$  for all  $t$ . We proceed by induction. Hence, assume that  $h_j(t) \geq \tilde{h}_j(t)$  for all  $t$  for some  $j \geq 1$ . This implies that  $T_j \leq \tilde{T}_j$ . For  $t < T_j$  we have by (??) that

$$\begin{aligned} \frac{d}{dt}(h_{j+1}(t) - \tilde{h}_{j+1}(t)) &= -\frac{1}{\tilde{h}_j(t)}(h_{j+1}(t) - \tilde{h}_{j+1}(t)) \\ &\quad + \left( \frac{1}{\tilde{h}_j(t)} - \frac{1}{h_j(t)} \right) h_{j+1}(t). \end{aligned}$$

Hence we see that  $h_{j+1}(t) \geq \tilde{h}_{j+1}(t)$  for all  $t < T_j$ . For  $t \in (T_j, \tilde{T}_j)$  the above equation still holds with  $h_j(t) = N$ ,

hence we still have  $h_{j+1}(t) \geq \tilde{h}_{j+1}(t)$ . For  $t > \tilde{T}_j$ ,  $t < \min\{T_{j+1}, \tilde{T}_{j+1}\}$  we have

$$\frac{d}{dt}(h_{j+1}(t) - \tilde{h}_{j+1}(t)) = -\frac{1}{N}(h_{j+1}(t) - \tilde{h}_{j+1}(t))$$

and we again have  $h_{j+1}(t) \geq \tilde{h}_{j+1}(t)$ . Hence  $T_{j+1} \leq \tilde{T}_{j+1}$  and  $h_{j+1}(t) \geq \tilde{h}_{j+1}(t)$  for all  $t$ .  $\blacksquare$

Given initial conditions the system (??) can be solved as follows. Define  $Q_0(t) = e^{\frac{t}{N}}$ , and for  $i > 1$  define

$$Q_i(t) = \int_0^t Q_{i-1}(z) dz + h_i(0)Q_{i-1}(0).$$

(Note that in general we have  $Q_i(0) = \prod_{j=1}^i h_j(0)$ .) The solution to (??) for  $t \leq T_1$  is then given by

$$h_i(t) = \frac{Q_i(t)}{Q_{i-1}(t)}$$

We can verify this directly: First note that the initial conditions are satisfied and that the solution is correct for  $i = 1$ . The key point is that  $\frac{d}{dt}Q_i(t) = Q_{i-1}(t)$ , so for  $i > 1$  we obtain

$$\begin{aligned} \frac{d}{dt}h_i(t) &= \frac{Q_{i-1}(t)}{Q_{i-1}(t)} - \frac{Q_i(t)Q_{i-2}(t)}{Q_{i-1}^2(t)} \\ &= 1 - \frac{h_i(t)}{h_{i-1}(t)}. \end{aligned}$$

Let us consider the system initialized with  $h_i(0) = 0$  for  $i \geq 1$ . For  $t < T_1$  we then have

$$Q_i(t) = N^i \sum_{k=i}^{\infty} \frac{(t/N)^k}{k!}$$

which yields

$$h_i(t) = \frac{t}{i} (1 + O(t/N)).$$

We now have two ways of analyzing the system, one using the differential equations and another using the algebraic approach based on the above. The differential equations are useful for establishing monotonicity properties of the solution and the algebraic approach is useful for characterizing limiting behavior.

We will first consider the analysis of the differential equations and focus primarily on the interval  $[0, T_1]$ . We assume initial conditions  $h_i(0)$  that are non-increasing in  $i$  and satisfy  $h_1(0) < 1$ . Let us denote  $T_1$  simply as  $T$ . Define for  $i \geq 1$ ,

$$r_i(t) = \frac{h_i(t)}{h_{i-1}(t)}; \quad R_i(t) = \frac{1}{2 - r_i(t)}; \quad \alpha_i(t) = \frac{2 - r_i(t)}{h_i(t)};$$

then we have for  $i > 1$ ,

$$\frac{d}{dt}r_i(t) = \alpha_{i-1}(t)(R_{i-1}(t) - r_i(t)) \quad (1)$$

Thus, we observe that  $r_i(t)$  tracks  $R_{i-1}(t)$ .

Let us call the sequence  $h_i$  *regularly ordered* if  $h_i$  is strictly positive, monotonically decreasing with  $h_1 \leq 1$ ,  $r_i \geq R_{i-1}$  for all  $i \geq 2$ .

**Lemma 7.** *Assume a regularly ordered initial condition  $h_i(0)$ . Assume further that initializing with  $h_i(0)$  gives  $r_2(T) \geq \frac{1}{2 - \frac{1}{N}}$ .*



Then the solution has  $r_i(t)$  non-increasing and  $r_i(t) \geq R_{i-1}(t)$  for all  $i \geq 2$  and  $t \in [0, T]$ .

*Proof:* We first remark that each  $\alpha_i$  is finite and lies in a finite interval bounded away from 0. The proof is by induction on  $i$ . First note that  $R_1(t) = \frac{1}{2 - \frac{1}{h_1(t)}}$  is non-decreasing and, if  $N < \infty$  it is increasing. It therefore follows from (1) that the assumed condition  $r_2(T) \geq \frac{1}{2 - \frac{1}{N}} = R_1(T)$  implies that  $r_2(t) \geq R_1(t)$  and that  $r_2(t)$  is non-increasing for all  $t \in [0, T]$ . We proceed by induction. Assume for some  $i \geq 2$  that  $r_i(t)$  is non-increasing and  $r_i(t) \geq R_{i-1}(t)$  for  $t \in [0, T]$ . Then  $R_i(t) = \frac{1}{2 - \frac{1}{r_i(t)}}$  is non-increasing. Since  $R_i(0) \leq r_{i+1}(0)$  we conclude from (1) that  $r_{i+1}(t)$  is non-increasing and  $r_{i+1}(t) \geq R_i(t)$  for  $t \in [0, T]$ . ■

Note that the above lemma implies that if an initial condition  $h_i(0)$  is regularly ordered and satisfies  $r_2(T) \geq R_1(T)$ , then  $h_i(t)$  is regularly ordered for all  $t \in [0, T]$ .

**Lemma 8.** Assume two regularly ordered initial conditions  $h'_i(0)$  and  $h_i(0)$  where  $h'_1(0) \geq h_1(0)$ , and  $r'_i(0) \geq r_i(0)$  for all  $i \geq 2$ . Assume further that  $r_2(T) \geq \frac{1}{2 - \frac{1}{N}}$ . Then  $h'_i(T') \geq h_i(T)$  and  $r'_i(T') \geq r_i(T)$  for all  $i \geq 1$ .

*Proof:* By Lemma 7, for  $i \geq 2$   $r_i(t)$  is non-increasing on  $[0, T]$  and  $r_i(t) \geq R_{i-1}(t)$ .

Note that  $T' \leq T$  and consider first  $t \in [0, T']$ . We have  $r'_1(t) \geq r_1(t)$  and  $h'_1(t) \geq h_1(t)$  since  $h'_1(0) \geq h_1(0)$ . We proceed by induction. Assume for some  $i \geq 1$  that  $r'_i(t) \geq r_i(t)$  and  $h'_i(t) \geq h_i(t)$ . Then  $R'_i(t) \geq R_i(t)$  and  $\alpha'_i(t) \leq \alpha_i(t)$ . From (1) we have

$$\begin{aligned} & \frac{d}{dt}(r'_{i+1}(t) - r_{i+1}(t)) \\ &= \alpha'_i(t)(R'_i(t) - r'_{i+1}(t)) - \alpha_i(t)(R_i(t) - r_{i+1}(t)) \\ &= -\alpha'_i(t)(r'_{i+1}(t) - r_{i+1}(t)) + \alpha'_i(t)(R'_i(t) - R_i(t)) \\ & \quad + (\alpha_i(t) - \alpha'_i(t))(r_{i+1}(t) - R_i(t)). \end{aligned}$$

By Lemma 7 the last term is non-negative and the second term is non-negative by the above argument. Since  $r'_{i+1}(0) - r_{i+1}(0) \geq 0$  the above equation clearly implies  $r'_{i+1}(t) - r_{i+1}(t) \geq 0$ . Since  $h'_i(t) \geq h_i(t)$  this implies  $h'_{i+1}(t) \geq h_{i+1}(t)$ . By induction we now have  $r'_i(t) \geq r_i(t)$  on  $[0, T']$  for all  $i \geq 1$ .

By Lemma 7,  $r_i(t)$  is decreasing on  $[T', T]$  and we conclude that  $r_i(T) \leq r_i(T') \leq r'_i(T')$ . Since  $h'_1(T') = h_1(T) = 1$  we obtain  $h'_i(T') = \prod_{j=2}^i r'_j(T') \geq \prod_{j=2}^i r_j(T) = h_i(T)$ . ■

We will now introduce some additional notation to capture the renewal nature of the system. Let us define  $h_i^{[k]}(t) = h_i(T_k + t)$  and  $T^{[k]} = T_k - T_{k-1}$ . Thus, we consider solving the system in a sequence of rounds and we use the superscript  $^{[k]}$  to denote round  $k$ . The initial condition for round  $k$  is taken from the ending state of round  $k-1$ . Hence  $h_i^{[k]}(0) = h_{i+1}^{[k-1]}(T^{[k-1]})$  for  $i \geq 1$ .

Let us call a sequence  $h_i(0)$  *fixed point convergent* if  $h_i(0)$  is regularly ordered and we have  $r_2(T) \geq \max\{R_1(T), h_1(0)\}$ , and  $r_i(T) \geq r_{i-1}(0)$  for  $i \geq 3$ .

**Lemma 9.** If  $h_i(0)$  is fixed point convergent then  $h_i^{[k]}(0)$  is fixed point convergent for each  $k$  and is monotonically increasing  $k$  converging to a fixed point.

*Proof:* The proof proceeds by induction. Assume  $h_i^{[k]}(0)$  is fixed point convergent. Let us consider the interval  $t \in [0, T^{[k]}]$ . By Lemma 7,

$$r_i^{[k]}(t) \geq R_{i-1}^{[k]}(t) \quad (2)$$

for all  $i \geq 2$ . Since  $\frac{1}{2-x} \geq x$  for  $x \in [0, 1]$ , we have  $R_{i-1}^{[k]}(t) \geq r_{i-1}^{[k]}(t)$  hence  $r_i^{[k]}(t) \geq r_{i-1}^{[k]}(t)$  for all  $i \geq 2$ .

We now consider  $h_i^{[k+1]}(0) = h_{i+1}^{[k]}(T)$  for  $i \geq 1$ . Let us first show  $h_i^{[k+1]}(0) \geq h_i^{[k]}(0)$ . We have

$$h_1^{[k+1]}(0) = h_2^{[k]}(T) = r_2^{[k]}(T) \geq h_1^{[k]}(0)$$

where the last step is by assumption that  $h_i^{[k]}$  is fixed point convergent. For  $i \geq 2$  we have

$$r_i^{[k+1]}(0) = r_{i+1}^{[k]}(T^{[k]}) \geq r_i^{[k]}(0)$$

where again, the last step is by assumption. We now obtain for  $i > 1$

$$\begin{aligned} h_i^{[k+1]}(0) &= h_1^{[k+1]}(0) \prod_{j=2}^i r_j^{[k+1]}(0) \\ &\geq h_1^{[k]}(0) \prod_{j=2}^i r_j^{[k]}(0) = h_i^{[k]}(0). \end{aligned}$$

Now we will show that  $h_i^{[k+1]}(0)$  is also fixed point convergent. Clearly, the sequence is monotonically decreasing and positive. For  $i \geq 3$  we have

$$r_i^{[k+1]}(0) = r_{i+1}^{[k]}(T) \geq R_i^{[k]}(T) = R_{i-1}^{[k+1]}(0)$$

where the middle inequality uses (2). Since we also have  $r_2^{[k+1]}(0) = r_3^{[k]}(T) \geq R_2^{[k]}(T) > R_1^{[k+1]}(0)$  we see that  $h_i^{[k+1]}(0)$  is regularly ordered.

Consider now  $r_2^{[k+1]}(T^{[k+1]})$ . Let  $\tau$  be determined by  $h_1^{[k]}(\tau) = h_1^{[k+1]}(0)$ . Then for  $t \leq T^{[k+1]} - \tau$  we have  $h_1^{[k]}(\tau + t) = h_1^{[k+1]}(t)$  hence  $\tau + T^{[k]} = T^{[k+1]}$  and  $R_1^{[k]}(\tau + t) = R_1^{[k+1]}(t)$ . Since  $r_2^{[k+1]}(0) \geq r_2^{[k]}(0) \geq r_2^{[k]}(\tau)$  we see that (1) now implies that  $r_2^{[k+1]}(t) \geq r_2^{[k]}(\tau + t)$ . Hence

$$r_2^{[k+1]}(T^{[k+1]}) \geq r_2^{[k]}(T^{[k]}) \geq \frac{1}{2 - \frac{1}{N}}$$

and since  $r_2^{[k]}(T^{[k]}) = h_1^{[k+1]}(0)$  we have

$$r_2^{[k+1]}(T^{[k+1]}) \geq \max\{h_1^{[k+1]}(0), \frac{1}{2 - \frac{1}{N}}\}.$$

We now apply Lemma 8 to obtain for  $i \geq 3$ ,

$$r_i^{[k+1]}(T^{[k+1]}) \geq r_i^{[k]}(T^{[k]}) = r_{i-1}^{[k+1]}(0).$$

We conclude that  $h_i^{[k]}$  is a monotonically increasing sequence that therefore has a limit. Correspondingly,  $T^{[k]}$  is a monotonically decreasing sequence that also has limit. The limit is a fixed point by continuity. ■

### D. Fixed Points

Simulations show that  $h_i^{[k]}(0)$  converges in  $k$  to a fixed point solution. In this section we solve for the set of fixed points.

In general for  $i \geq 1$  we have  $h_i^{[k]}(0) = \frac{Q_{i+1}^{[k-1]}(T^{[k-1]})}{Q_i^{[k-1]}(T^{[k-1]})}$ . Since  $Q_0^{[k]}(0) = 1$  for all  $k$  we have  $Q_1^{[k]}(0) = h_1^{[k]}(0) = e^{-\frac{T^{[k-1]}}{N}} Q_2^{[k-1]}(T^{[k-1]})$ . Since  $Q_i^{[k]}(0) = h_i^{[k]}(0) Q_{i-1}^{[k]}(0)$  we proceed inductively to obtain

$$Q_i^{[k]}(0) = e^{-\frac{T^{[k-1]}}{N}} Q_{i+1}^{[k-1]}(T^{[k-1]}).$$

Hence, we can express  $Q^{[k]}(t)$  directly in terms of  $Q^{[k-1]}(T^{[k-1]})$  as follows.

$$\begin{aligned} Q_i^{[k]}(t) &= e^{-\frac{T^{[k-1]}}{N}} \sum_{j=0}^{i-1} Q_{i+1-j}^{[k-1]}(T^{[k-1]}) \frac{1}{j!} t^j + N^i \sum_{j=i}^{\infty} \frac{(t/N)^j}{j!} \\ &= \sum_{j=0}^{i-1} Q_{i-j}^{[k]}(0) \frac{1}{j!} t^j + N^i \sum_{j=i}^{\infty} \frac{(t/N)^j}{j!}. \end{aligned}$$

Noting at the fixed point we have  $Q_i(T) = e^{\frac{T}{N}} Q_{i-1}(0)$  we obtain

$$\begin{aligned} e^{\frac{T}{N}} Q_{i-1}(0) &= \sum_{j=0}^{i-1} Q_{i-j}(0) \frac{1}{j!} T^j + N^i \sum_{j=i}^{\infty} \frac{(T/N)^j}{j!} \\ &= \sum_{j=0}^i Q_{i-j}(0) \frac{1}{j!} T^j + N^i \sum_{j=i+1}^{\infty} \frac{(T/N)^j}{j!} \end{aligned}$$

Dividing through by  $Q_{i-1}(0)$  and rearranging terms we obtain

$$\begin{aligned} h_i(0) &= e^{\frac{T}{N}} - T - T \sum_{j=2}^i \frac{1}{j!} \prod_{k=1}^{j-1} \frac{T}{h_{i-k}(0)} \\ &\quad - \frac{1}{Q_{i-1}(0)} N^i \sum_{j=i+1}^{\infty} \frac{(T/N)^j}{j!} \end{aligned}$$

Define  $h_i(0) = N$  for  $i \leq 0$  we can write this as

$$h_i(0) = e^{\frac{T}{N}} - T - T \sum_{j=2}^{\infty} \frac{1}{j!} \prod_{k=1}^{j-1} \frac{T}{h_{i-k}(0)}$$

Note that this confirms  $h_1(0) = e^{\frac{T}{N}} - N(e^{\frac{T}{N}} - 1)$ . It follows that  $h_i(0)$  is a monotonically decreasing in  $i$  and therefore has a limit  $h_{\infty}$ . For  $T = 0$  we obtain  $h_i(0) = 1$  for all  $i$  and if  $T > 0$  then we must have  $h_{\infty} > 0$ . Taking limits we obtain

$$h_{\infty} = e^{\frac{T}{N}} - T - T \sum_{j=2}^{\infty} \frac{1}{j!} \prod_{k=1}^{j-1} \frac{T}{h_{\infty}} = e^{\frac{T}{N}} - h_{\infty}(e^{\frac{T}{N}} - 1)$$

or

$$h_{\infty} = e^{-T(\frac{1}{h_{\infty}} - \frac{1}{N})}, \quad T = -\frac{h_{\infty} \ln h_{\infty}}{1 - \frac{h_{\infty}}{N}}. \quad (3)$$

Now,  $-\frac{h_{\infty} \ln h_{\infty}}{1 - \frac{h_{\infty}}{N}}$  is a concave function of  $h_{\infty}$  on  $[0, 1]$ . Let  $T^F(N)$  denote its maximum. We have  $T^F(\infty) = e^{-1}$ .

**Lemma 10.** For each  $T \leq T^F(N)$  there exists a fixed point with the given  $T$ .

*Proof:* We need only show that the recursion for  $h_i(0)$  is well behaved. Let  $T$  be as above and assume  $h_j(0) \geq h_{\infty} = h_{\infty}(T)$ . Then we have

$$\begin{aligned} h_i(0) &= e^{\frac{T}{N}} - T - T \sum_{j=2}^{\infty} \frac{1}{j!} \prod_{k=1}^{j-1} \frac{T}{h_{i-k}(0)} \\ &\geq e^{\frac{T}{N}} - T - T \sum_{j=2}^{\infty} \frac{1}{j!} \prod_{k=1}^{j-1} \frac{T}{h_{\infty}} \\ &\geq e^{\frac{T}{N}} - h_{\infty}(e^{\frac{T}{N}} - 1) \\ &= h_{\infty} \end{aligned}$$

Since  $h_i = N$  for  $i \leq 0$  we see that the sequence is well defined. ■

Let us denote the fixed point corresponding to  $T^F(N)$  as  $h^F = h^F(N)$ . Since  $h_{\infty}^F(N) = \operatorname{argmax}_{1-\frac{h}{N}} \frac{-h \ln h}{1-\frac{h}{N}}$  a little calculus shows that  $h_{\infty}^F(N)$  is determined by  $-\ln h_{\infty}^F(N) = 1 - \frac{h_{\infty}^F(N)}{N}$  or  $eh_{\infty}^F(N) = e^{\frac{h_{\infty}^F(N)}{N}} \geq 1 + \frac{h_{\infty}^F(N)}{N}$  and we obtain the bound

$$h_{\infty}^F(N) \geq \frac{N}{Ne - 1}. \quad (4)$$

Moreover we have from (3)

$$T^F(N) = \frac{-h_{\infty}^F(N) \ln h_{\infty}^F(N)}{1 - \frac{h_{\infty}^F(N)}{N}} = h_{\infty}^F(N). \quad (5)$$

From (3) we obtain

$$h_1^F = h_{\infty}^F e^{T^F/h_{\infty}^F} - h_{\infty}^F = (e - 1)h_{\infty}^F \quad (6)$$

By Lemma 6 the fixed point  $h^F(N)$  gives an upper bound on the solution for  $N$ . This shows that the asymptotic gain in the transmission velocity of the file is upper bounded by that from  $h^F(N)$ . In the case  $N = \infty$  we prove in the next section that the solution converges to  $h^F(\infty)$ . Simulations indicate that convergence to  $h^F(N)$  occurs for all  $N > 1$  but we do not have a proof for the general case. We can show, however, that the asymptotic acceleration is that determined by the fixed point.

### E. Convergence

The case  $N = \infty$  is the Fountain code case (infinitely low rate). Let us initialize in round 1 with  $h_i^{[1]}(0) = 0$  for  $i > 0$ . It is immediate that

$$Q_i^{[1]}(t) = \frac{1}{i!} t^i; \quad h_i^{[1]}(t) = \frac{t}{i}; \quad r_i^{[1]}(t) = \frac{i-1}{i}.$$

Hence, the case  $k = 1$  is fixed point convergent in that  $h_i^{[1]}(t) = \frac{t}{i}$  and  $r_i(t) = R_{i-1}(t)$  for all  $t$ . Hence the solution in round 1 is fixed point convergent.

**Lemma 11.** In the  $N = \infty$  case the sequence  $h^{[k]}$  is monotonically increasing and approaches the limit  $h^F$ .

*Proof:* Since the solution above in round 1 is fixed point convergent in the  $N = \infty$  case, convergence of  $h^{[k]}$  follows directly from Lemma 9. Since  $T^{[k]}$  is decreasing it has a limit and it follows from Lemma 8 that this limit must be  $T^F(\infty)$ .

Indeed, if the limit were less than  $T^F(\infty)$  then we would obtain a contraction to Lemma 8 by initializing the system with  $h^F(\infty)$  and comparing the 0 initialization. Hence the limit of  $h^{[k]}$  is  $h^F(\infty)$ . ■

In the case of finite  $N$  we see that  $r_i^{[1]}(t)$  is increasing and so the above monotonicity argument does not succeed. Simulations indicate that in subsequent rounds the sequence does become fixed point convergent, but we have not been able to prove this. Thus we are unable to prove convergence of  $h^{[k]}$  although we conjecture that it converges to  $h^F(N)$ .

The upper bound in Theorem 2 is obtained by initializing with the fixed point  $h^F(N)$ , which obviously has the stated asymptotic delay. The lower bound is obtained by showing that the solution eventually exceeds a fixed point convergent condition. Then Lemma 9 provides the lower bound.

Before giving the proof for the lower bound we develop some more preliminary results.

**Lemma 12.** *For a fixed initial condition we can parameterize the solution on  $[0, T]$  by  $h_1$ . Then  $h_i(h_1; N)$  is a decreasing function of  $N$ .*

*Proof:* We consider parameterizing the solution on  $[0, T]$  by  $h_1$  which spans the interval  $[h_1(0), 1]$ . Then we have  $h_2(h_1)$  satisfies

$$\frac{d}{dh_1} h_2 = \left(1 - \frac{h_2}{h_1}\right) \left(\frac{1}{1 - \frac{h_1}{N}}\right)$$

from which it easily follows that  $h_2(h_1)$  is a decreasing function of  $N$ . Now assume that  $h_{i-1}(h_1)$  is a decreasing function of  $N$ . Since we have

$$\frac{d}{dh_1} h_i = \left(1 - \frac{h_i}{h_{i-1}}\right) \left(\frac{1}{1 - \frac{h_1}{N}}\right)$$

it is easily seen that  $h_i(h_1)$  is a decreasing function of  $N$ . Hence the lemma follows by induction. ■

**Lemma 13.** *For any  $N \geq N' > 1$  we have  $h(\gamma t; N') \geq h(t; N)$  where  $\gamma = \frac{1 - \frac{1}{N}}{1 - \frac{1}{N'}}$ .*

*Proof:* By Lemma 12 we see that for round [1] we have  $h_i(h_1; N') \geq h_i(h_1; N)$  where we have parameterized by  $h_1$  instead of  $t$  and indicated explicit dependence on  $N$ .

Let us define  $\gamma(x) = \frac{1 - \frac{x}{N}}{1 - \frac{x}{N'}}$  so the above  $\gamma$  is  $\gamma(1)$ . Note that  $\gamma(x)$  is increasing in  $x$  on  $[0, 1]$ . Now consider a fixed initial condition. Then for  $\gamma t \leq T(N')$  and  $t \leq T(N)$  we

have

$$\begin{aligned} \frac{d}{dt} h_1(\gamma t; N') &= \gamma \left(1 - \frac{h_1(\gamma t; N')}{N'}\right) \\ &= \gamma (h_1(\gamma t; N')) \left(1 - \frac{h_1(\gamma t; N')}{N'}\right) \\ &\quad + (\gamma(1) - \gamma(h_1(\gamma t; N')))) \left(1 - \frac{h_1(\gamma t; N')}{N'}\right) \\ &= \left(1 - \frac{h_1(\gamma t; N')}{N}\right) \\ &\quad + (\gamma(1) - \gamma(h_1(\gamma t; N')))) \left(1 - \frac{h_1(\gamma t; N')}{N'}\right) \end{aligned}$$

and we obtain

$$\begin{aligned} \frac{d}{dt} (h_1(\gamma t; N') - h_1(t; N)) &= -\frac{1}{N} (h_1(\gamma t; N') - h_1(t; N)) \\ &\quad + (\gamma(1) - \gamma(h_1(\gamma t; N')))) \left(1 - \frac{h_1(\gamma t; N')}{N'}\right) \end{aligned}$$

hence  $h_1(\gamma t; N') \geq h_1(t; N)$ . It follows that  $T(N) \leq \gamma^{-1} T(N')$ . By Lemma 12 we have  $h_i(\gamma t; N') \geq h_i(t; N)$  for all  $i$  and also  $h(T(N'); N') \geq h(T(N), N)$ . We can now conclude that  $h(\gamma t; N') \geq h(t; N)$  for all  $t \in [0, T(N)]$ .

Now consider the system initialized with  $h = 0$ . By the above we have  $h(\gamma t; N') \geq h(t; N)$  for  $t \leq T_1(N)$ . If we reduce  $h(\gamma T_1(N); N')$  to set it equal to  $h(T_1(N); N)$  then the above argument would again yield  $h(\gamma t; N') \geq h(t; N)$  for  $t \in [T_1(N), T_2(N)]$ . Now, Lemma 6 implies that without the reduction  $h(\gamma t; N')$  would be larger still so we have  $h(\gamma t; N') \geq h(t; N)$  for  $t \in [T_1(N), T_2(N)]$  for the actual solution. The same argument can be repeated for  $t \in [T_i(N), T_{i+1}(N)]$  for  $i = 2, 3, \dots$  and hence by induction we obtain  $h(\gamma t; N') \geq h(t; N)$  for all  $t$ . ■

Let us define for  $i \geq 1$  and  $N \in (1, \infty]$ ,

$$h_i^*(N) = \frac{1}{(i+1) - \frac{i}{N}} \text{ and } r_i^*(N) = \frac{h_i^*}{h_{i-1}^*}$$

with  $h_0^*(N) = N$ . The key property of this definition is that for  $i \geq 2$  we have

$$r_i^*(N) = \frac{1}{2 - r_{i-1}^*(N)} \quad (7)$$

**Lemma 14.** *The initial condition  $h^*(M)$  is fixed point convergent for a given  $N \leq M$  if, given the initial condition, we have  $h_2(T) \geq h_1^*(N)$ . Moreover, we then have  $r_i(T) \geq r_{i-1}^*(N)$ , for all  $i \geq 3$ , hence  $h_i(T) \geq h_{i-1}^*(N)$ , for all  $i \geq 3$ .*

*Proof:* It follows from (7) that  $h_i^*(M)$  is regularly ordered. To prove it is fixed point convergent for  $N$  we need to show that  $r_2(T) \geq \max\{R_1(T), h_1(0)\}$ , and that  $r_i(T) \geq r_{i-1}(0)$  for  $i \geq 3$ . We assume that  $r_2(T) = h_2(T) \geq \frac{1}{2 - \frac{1}{N}} = R_1(T) = h_1(0)$  so to prove it is fixed point convergent for  $N$  we need only show that  $r_i(T) \geq r_{i-1}(0)$  for  $i \geq 3$ . Since  $r_{i-1}(0) = r_{i-1}^*(M) \leq r_{i-1}^*(N)$  we see that it is sufficient to show  $r_i(T) \geq r_{i-1}^*(N)$  for all  $i \geq 3$ .

We have by assumption that  $r_2(T) \geq r_1^*(N)$ . We proceed by induction. Hence assume that  $r_i(T) \geq r_{i-1}^*(N)$ . It follows from Lemma 7 that  $r_i(t)$  is decreasing and  $r_i(t) \geq R_{i-1}(t)$  for all  $i \geq 2$ . Hence  $r_{i+1}(T) \geq R_i(T) = \frac{1}{2-r_i(T)} \geq \frac{1}{2-r_{i-1}^*(N)} = r_i^*(N)$ . ■

**Lemma 15.** *There exists  $\delta > 0$  such that if  $N \in [2, 6]$  then the initial condition  $h_i^*(M)$  is fixed point convergent for  $N$  for all  $M \in [N, N + \delta]$ . If  $N \geq 6$  then the initial condition  $h_i^*(\infty)$  is fixed point convergent.*

*Proof:* By Lemma 14  $h_i^*(M)$  is fixed point convergent for  $N$  if  $h_2(T) \geq h_1^*(N)$ . We can obtain conditions for this using the solution obtained in terms of the functions  $Q_i(t)$ . In particular we have

$$\begin{aligned} Q_0(T) &= e^{\frac{T}{N}} \\ Q_1(T) &= h_1^*(M) + N(e^{\frac{T}{N}} - 1) \\ Q_2(T) &= h_1^*(M)h_2^*(M) + h_1^*(M)T + N^2(e^{\frac{T}{N}} - 1 - \frac{T}{N}) \end{aligned}$$

and  $T = T(N)$  is determined by  $Q_1(T) = Q_0(T)$ . We have  $h_2(T) = Q_2(T)/Q_1(T)$  so the initial condition is fixed point convergent if

$$Q_2(T) \geq Q_1(T)h_1^*(N)$$

under the condition  $Q_1(T) = Q_0(T)$ , which determines  $T$ .

The condition  $Q_1(T) = Q_0(T)$  gives

$$Q_0(T) = Q_1(T) = e^{\frac{T}{N}} = \frac{N - h_1^*(M)}{N - 1} = 1 + \frac{1 - h_1^*(M)}{N - 1}$$

hence

$$T = N \ln\left(1 + \frac{1 - h_1^*(M)}{N - 1}\right)$$

Let us first consider the case  $M = \infty$ . We have  $h_1^*(M) = 1/2$  and  $h_2^*(M) = 1/3$  and  $e^{\frac{T}{N}} = \frac{2N-1}{2(N-1)}$ . Using  $N^2(e^{\frac{T}{N}} - 1 - \frac{T}{N}) \geq \frac{1}{2}T^2$  we see that  $h_i^*(\infty)$  is fixed point convergent for  $N$  if

$$\frac{1}{6} + \frac{1}{2}T + \frac{1}{2}T^2 \geq \frac{N}{2(N-1)} = \frac{1}{2} + \frac{1}{2N-2}$$

Some algebra shows that  $T(N) + T(N)^2 - \frac{1}{N-1}$  is increasing in  $N$  on  $[2, \infty)$  (although we only use this for  $N \geq 6$ ) and we see that if the equality holds for  $N = N'$  then it holds for all  $N \geq N'$ . It can be easily verified that the inequality holds for  $N = 6$ .

It is clear that  $h_1(0)$  and  $h_2(0)$  are uniformly continuous in  $M$  on  $[2, 6]$ . Thus, to obtain the desired result we need only verify that  $Q_2(T) > e^{\frac{T}{N}} \frac{1}{2-N}$  when  $M = N$  for  $N \in [2, 6]$ . Noting that for  $N = M$  we have  $h_1(0) = \frac{1}{2-N}$  we can, by dividing through by  $h_1(0)$  and rearranging terms, write the condition to be shown as

$$\left(1 - \frac{1}{N}\right)T + \left(2 - \frac{1}{N} - \frac{1}{N^2}\right) \sum_{k=2}^{\infty} \frac{1}{N^{k-2}} \frac{T^k}{k!} > \frac{2N-2}{3N-2}$$

where in this case we have  $T(N) = N \ln(1 + \frac{1}{2N-1})$ . This inequality holds on  $[2, 6]$  taking only the  $k = 2$  and  $k = 3$

terms from the sum, which can be verified with some algebra using the bound for  $N \geq 2$ ,

$$T(N) \geq N \left( \frac{1}{2N-1} - \frac{1}{2(2N-1)^2} + \frac{1}{4(2N-1)^3} \right).$$

*Main Proof of Theorem 2 - lower bound:* We assume  $N \geq 2$ .

We can obtain a lower bound by showing that for some  $t = \tau$  we have  $h(\tau; N) \geq \hat{h}$  where  $\hat{h}$  is fixed point convergent (for the given  $N$ .) To see why this produces a lower bound with the desired property consider initializing the system with  $\hat{h}$ . By Lemma 9 the resulting solution  $\hat{h}(t; N)$  then satisfies  $\lim_{k \rightarrow \infty} \hat{h}^{[k]}(N) = h^F(N)$  with  $\hat{T}^{[k]}$  a decreasing sequence that approaches  $T^F(N)$ . It follows that for any  $\epsilon > 0$  we have  $\hat{T}_k \leq k(T^F(N) + \epsilon)$  for  $k$  large enough. Now, by Lemma ??,  $\hat{h}(t; N) \leq h(\tau + t; N)$  so  $T_k \leq \tau + \hat{T}_k$ . Since  $\epsilon$  is arbitrary we now obtain  $\limsup_{k \rightarrow \infty} \frac{1}{k}T_k \leq T^F(N)$  which together with the lower bound gives the result.

By Lemma 13 we see that it is sufficient to find  $\hat{h}(t, N')$  that is fixed point convergent for  $N$  for any  $N' \geq N$ . For  $N \geq 6$  we have  $h_i(T_1(\infty); \infty) = \frac{1}{i}$  is fixed point convergent by Lemma 15. Since  $h_i(T_1(N); N) \leq h_i(T_1(\infty); \infty)$  for  $N < \infty$  by Lemma 12 the proof is complete for  $N \geq 6$ .

From Lemma 14 it follows that for some  $\tau$  we have  $h_i(\tau; N = 6) \geq h_i^*(6)$ . This gives a fixed point convergent sequence for  $N \in [6 - \delta, 6]$ . Similarly, it now follows from Lemma 14 that for some  $\tau$  we have  $h_i(\tau; N = 6 - \delta) \geq h_i^*(6 - \delta)$ . This gives a fixed point convergent sequence for  $N \in [6 - 2\delta, 6 - \delta]$ . Proceeding by induction we obtain fixed point convergent conditions for all  $N \geq 2$ .

We conclude by noting that the lower bound and upper bound are asymptotically equal. ■

### F. Sketch of Proof of Proposition??

As before there occurs a sequence of times  $T_1, T_2, \dots$  where  $T_i$  denotes the time  $t$  where  $h_i(t)$  reaches 1. For  $t \leq T_1$  we have

$$h_1(t) = h_1(0) + t$$

We obtain for  $i > 1$ ,

$$e^t h_i(t) = \int_0^t e^s h_{i-1}(s) ds + h_i(0). \quad (8)$$

Defining  $r_i := \frac{h_i}{h_{i-1}}$  for  $i \geq 2$  we get

$$\frac{d}{dt} r_i(t) = 1 - \frac{r_i(t)}{r_{i-1}(t)}. \quad (9)$$

where we also define  $r_1(t) := \frac{h_1(t)}{h_1(t)+1}$ .

Let us consider the initialization  $h_i(0) = 0, i \geq 1$  and the interval  $[0, T_1]$ . It is easy to see from 8 that  $r_i(t)$  approaches 0 as  $t \rightarrow 0$ .

**Lemma 16.** *If  $r_i(0) \leq \frac{1}{2}$  for  $i \geq 2$  then we have  $r_i(t) \leq \frac{1}{2}$  for  $i \geq 1$  and  $t \in [0, T]$ .*

*Proof:* Consider any initial condition  $r_i(0) \leq \frac{1}{2}$ . Then on  $[0, T]$  we have  $r_1(t) \leq \frac{1}{2}$  and so (9) implies  $r_i(t) \leq \frac{1}{2}$  for  $i = 2, 3, \dots$  by induction. ■

The above Lemma shows that  $h_i$  decays rapidly in  $i$ . This also implies that the probability that a transmission will be successful is at least  $\frac{1}{2}$  so the ratio of wasted to useful transmissions is finite.

Given an initial condition we can obtain a solution as

$$h_1(t) = h_1(0) + t$$

and for  $i \geq 1$ ,

$$e^t h_i(t) = \int_0^t e^s h_{i-1}(s) ds + h_i(0).$$

Even though we have a simple recursive form it appears difficult to prove convergence. Simulations indicate that  $h^{[k]}$  is an increasing sequence that therefore converges.

### G. Fixed Points

Consider  $M_k(t) = \sum_{i=k}^{\infty} h_i(t)$ . Since  $h_i(t)$  goes to 0 in  $i$  we have for  $k \geq 2$ ,

$$\frac{d}{dt} M_k = h_{k-1}$$

and at a fixed point we have

$$\int_0^T h_{k-1}(s) ds = M_k(T) - M_k(0) = M_{k-1}(0) - M_k(0) = h_{k-1}(0).$$

In particular setting  $k = 2$  we obtain gives the necessary condition

$$\int_0^T h_1(s) ds = h_1(0)T + \frac{1}{2}T^2 = h_1(0)$$

and since  $h_1(0) = 1 - T$  we can solve to obtain  $T = \sqrt{2}$ . Since this determines  $h_1(t)$  we can then in principle solve for  $h_i(0), i = 2, 3, \dots$  using

$$\int_0^T h_i(s) ds = h_i(0)$$

and (8).

Analysis of the long term evolution of  $M_1^{[k]}(0)$  shows that the asymptotic speed up is at least  $5/3$ .

In general we have

$$\begin{aligned} M_1^{[k+1]}(0) &= M_1^{[k]}(T^{[k]}) - 1 \\ &= M_1^{[k]}(0) + \int_0^{T^{[k]}} h_1^{[k]}(s) ds + (1 - h_1^{[k]}(0)) - 1 \\ &= M_1^{[k]}(0) + 2T^{[k]} - \frac{1}{2}(T^{[k]})^2 - 1 \end{aligned}$$

Let us define  $p(T) = 2T - \frac{1}{2}(T)^2 - 1$ . Since  $0 \leq M_1^{[k]}(0) \leq 1$  we easily obtain  $|\sum_{k=1}^i p(T^{[k]})| \leq 1$ . We have  $T^{[k]} \geq \frac{1}{2}$ . Since  $p(T) = 2T - \frac{1}{2}(T)^2 - 1$  is concave increasing on  $[\frac{1}{2}, 1]$  we see that for any probability distribution of  $T$  on  $[\frac{1}{2}, 1]$  the point  $(\mathbb{E}(T), \mathbb{E}(P(T)))$  is below the graph of  $p(T)$  and above the line segment joining the endpoints. Since  $p(\frac{1}{2}) = -\frac{1}{8}$  and

$p(1) = \frac{1}{2}$  any distribution of  $T$  with  $\mathbb{E}(P(T)) = 0$  has  $\mathbb{E}(T) \leq \frac{3}{5}$ . Hence,

$$\limsup_{i \rightarrow \infty} \frac{1}{i} T_i = \limsup_{i \rightarrow \infty} \frac{1}{i} \sum_{k=1}^i T^{[k]} \leq \frac{3}{5}.$$

We also have the corresponding lower bound (which we conjecture is tight) of

$$\liminf_{i \rightarrow \infty} \frac{1}{i} \sum_{k=1}^i T^{[k]} \geq 2 - \sqrt{2} \simeq 0.5857...$$